UCSF

Title
Origins of life: first came evolutionary dynamics.

Permalink

Authors
Kocher, Charles
Dill, Ken

Publication Date
2023

DOI
10.1017/qrd.2023.2

Peer reviewed

CAMBRIDGE
UNIVERSITY PRESS

Origins of life: first came evolutionary dynamics

Charles Kocher[1,2] and Ken A. Dill[1,2,3]

[1]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY, USA; [2]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA and [3]Department of Chemistry, Stony Brook University, Stony Brook, NY, USA

Abstract

When life arose from prebiotic molecules 3.5 billion years ago, what came first? Informational molecules (RNA, DNA), functional ones (proteins), or something else? We argue here for a different logic: rather than seeking a *molecule type*, we seek a *dynamical process.* Biology required an ability to evolve before it could choose and optimise materials. We hypothesise that the *evolution process* was rooted in the *peptide folding process.* Modelling shows how short random peptides can collapse in water and catalyse the elongation of others, powering both increased folding stability and emergent autocatalysis through a disorder-to-order process.

Requirement for life's origin: persistent propagation

What was the origin of life? A pre-requisite for answering that question is to define the difference between dead and alive. Defining life has been notoriously challenging (Schrodinger, 1944; Cleland and Chyba, 2002; Popa, 2004; Benner, 2010; Machery, 2012; Pross, 2016; Plaxco and Gross, 2021). For example, the ability to metabolise, grow, and duplicate is not sufficient to distinguish life from a candle flame. Nevertheless, a good consensus definition is from NASA: 'Life is a self-sustaining chemical system *capable of Darwinian evolution*' (Joyce *et al.*, 1994). The italics are ours; the clear implication is that life could not begin before some dynamical adaptation process of molecules was already operative.

So, in order to seek the origins of life, we ask what physico-chemical process(es) could have driven prebiotic molecules to become autocatalytic and adaptive. We call this *The Day Two Problem*, to distinguish it from traditional questions about 'What Came First', which we call *The Day One Problem.*

- *The 'Day One' Question:* What material came first? Think of the metaphorical 'chicken or the egg problem' (although in the real world, the chicken-and-egg framing is misleading because many things, including chickens and eggs, emerged in parallel, not in series). For example: Did life start as an RNA World (Gilbert, 1986; Joyce and Szostak, 2018)? Or a lipid world (Segré *et al.,* 2001; Deamer, 2011), an amyloid world (Maury, 2009, 2015, 2018) or through metabolism first (Wächtershäuser, 1988; De Duve and De Neufville, 1991; De Duve, 1995; Cody, 2004; Shapiro, 2006; Jordan *et al.,* 2021; Matsuo and Kurihara, 2021)? Our view is that the full origin of life required many aspects – metabolism, information, protein-like catalysts and makers, and others – to arise together.
- *The 'Day Two' Question:* What dynamical process might have driven prebiotic molecules to become self-sustaining and self-serving? How might prebiotic molecules undergo directed change from Day One to Day Two and beyond? For biology to *operate*, it requires an *operating system.* What would drive molecules to evolve further on their own?

We give here a hypothesis. We describe a mechanism by which prebiotic undirected syntheses of short peptides could plausibly become 'makers', that is molecules that persistently make other molecules. We give a summary and perspective of recent computer modelling of a disorder-to-order process that achieves positive feedback against the forces of degradation. We start below by positing Darwinian evolutionary dynamics as a driven machine cycle of steps, because this vantage point illuminates principles about possible molecular origins.

Description of Darwinian dynamics as a cyclic machine

Darwinian Dynamics has three well-known features: (1) Replication, moms making more moms; (2) Mutation, search and discovery of sequence → function polymers that create new molecular entities and mechanisms and (3) Selection, competition-driven upratcheting of fitness. Fig. 1 expresses these properties in the form of a machine-like, biosphere-wide nonequilibrium (NEQ) cycle that we call the *Darwinian Evolution Machine* (DEM). Described in detail elsewhere (Kocher and Dill, 2023), the cycle operates from left to right. (*a*) At time *t*, *X* indicates a wild-type (status quo) population. (*b*) A mutation occurs in some individual. (*c*) That individual cell is
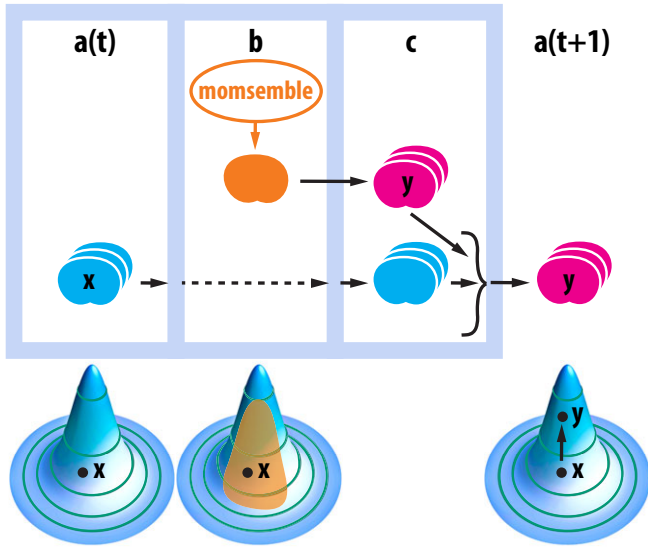
Fig. 1. Top: The DEM cycle. (*a*) At time *t* shows a population *X* of wild-type cells. (*b*) One cell mutates. That cell grows to have population *Y* in (*c*). Populations *X* and *Y* compete, one wins, and the cycle begins again as (*a*) at time *t* + 1. Bottom: Fitness landscapes show the separation of actions. From point *X*, mutation entails a random, relatively unbiased exploration (orange region). The third landscape shows selection, in this case of *Y*, where the bias and preference occurs.

grown up into a population Y. Populations X and Y compete for resources. ($a(t+1)$) The winner becomes the new status quo wild-type population, thus becoming 'remembered' in the population, and gains more resources. The cycle repeats, driven by a persistent external supply of resources.

The steps (*b*) and (*c*) of population growth on the resources, competition and winning can be expressed as population-resource dynamics. For example, if we have N competitors A_n, labelled by their phenotype, fighting for one resource r,

$$\frac{dr}{dt} = g(r) - D_r(r) - \sum_n U_n(A_n, r),$$

$$\frac{dA_n}{dt} = R_n\left(\frac{U_n(A_n, r)}{A_n}\right)A_n - D_n(A_n),$$

(1)

where $g(r)$ describes the NEQ input of the resource, $D_r(r)$ and $D_n(A_n)$ are decay/death terms, $U_n(A_n, r)$ is the total resource use rate by moms of type n, and R_n is the reproduction rate of one mom of type n given that mom eats resource at a rate $U_n(A_n, r)/A_n$. Mutational discovery will introduce new moms, say A_{N+1}, which are then selected for or against by the DEM. The most competitive moms are 'remembered' by the DEM because they are good enough autocatalysts to maintain persistent populations. 'Fitness', which is the term that describes this competition-driven selection, is non-trivial to define, and can be model-dependent; see Supplementary Material SI.2 for more discussion.

Known biology constrains the mathematical form that is needed for the function U. Often, population genetics modelling approximates it as linear in both resources and moms, $U_n(A_n, r) = rk_nA_n$. But, such linearity leads to 'winner-take-all' (WTA) dynamics (Volterra, 1928; Fisher, 1930; Gause, 1934; MacArthur, 1970; Hsu *et al.,* 1977; Tilman, 1982; Chesson, 1990; Lifson, 1997; Pross, 2011; van Opheusden *et al.,* 2015), a dynamics that misses important features of evolution. Instead, evolution often gives 'peaceful coexistence' of multiple species on a given resource (Hutchinson, 1961; Armstrong and McGehee, 1980; Chesson, 2000; Chesson and

Kuang, 2008; Charlebois and Balázsi, 2016; Barabás *et al.,* 2018; Goyal *et al.,* 2018; Wang *et al.,* 2022). Peaceful coexistence is captured using a saturating function (Beddington, 1975; DeAngelis *et al.,* 1975; Novak and Stouffer, 2021; Stouffer and Novak, 2021; Kocher and Dill, 2023),

$$U_n(A_n, r) = \frac{rk_nA_n}{b_n + c_nr + A_n},$$

(2)

which simply expresses two natural limits, that maximum concentrations of moms are finite and that speeds of producing offspring are finite. What is novel in the present DEM perspective is the combining of this generalised form of population-genetics (Eq. (1)) with iterative mutation, competition and selection cycles (Kocher and Dill, 2023). In the following section, we extract four principles from this DEM perspective that helps us formulate possible molecular precursors in the next sections.

Features of evolution that are relevant for origins

The DEM model perspective illuminates what is needed for the origin of life. First, the DEM is a *maker of makers*, a process of moms creating more moms. The DEM is an *autocatalytic set, or a collection of entities, each of which can be created catalytically by other entities, such that as a whole, the set is able to catalyse its own production* (Hordijk, 2019). An extensive literature describes the importance of autocatalytic sets in the origins of life (Eigen, 1971; Kauffman, 1971, 1986; Kauffman *et al.,* 1993; Dyson, 1999, 1982; Jain and Krishna, 2002; Hordijk and Steel, 2014; Hordijk, 2019; Hordijk *et al.,* 2022). The positive feedback of makers making makers contributes to the self-sustaining nature of evolution.

Second, environments that are unruly and fluctuating can sort winners from losers through booms and busts (Doebeli *et al.,* 2021; Wang *et al.,* 2022). Booms and busts drive the recycling of resources, taking resources away from the losers and giving them to the winners, thus driving the rich to get richer on the road to autocatalysis.

Third, DEM Dynamics can sustain *peaceful coexistence* among multiple agents at the same time. In a world of winners-taking-all (WTA), without peaceful coexistence, evolution would have been brittle, always on the edge of extinction. If X is more fit than Y in environment E_1, X would be the lone survivor in a WTA model. Now, if the environment fluctuates to E_2, which kills X, then the whole ecosystem dies. Instead, in a world of coexistence, diversity preserves the ecosystem. An ecosystem that has an ensemble of *backup moms* is more robust to unpredictable new environments. Ensembles are crucial for long-term survival and persistence of the DEM.

And fourth, the biosphere-wide DEM is a *driven* machine: its cycles of molecule-making are powered by uptake of out-of-equilibrium resources from the environment. There are different tendencies for driven systems than for equilibrium processes. Some detail is given in Supplementary Material SI.1; here we just give a few examples. (1) A fluid subject to gravity will flow down a hill and stop in the valley at the bottom. But, a fluid subject to a strong force can flow beyond the valley to cross the next hill and beyond. (2) A TV set or computer performs intricate functions as long as it is 'plugged in'. Its current flows are not predicted by principles of equilibrium, such as the Second Law. Such devices only tend to equilibrium when they are unplugged. Think about an electromagnet. A metal rod will not pick up nails, but when a current is driven around the bar, it will. An electromagnet is driven by the current

input. Its action is a nonequilibrium (NEQ) force. That force goes to zero when the input current is turned off. (3) While equilibrium systems tend downhill in energy (or free energy), driven systems can also go uphill. Think of chemical reactions, like binding events or protein folding or molecular association or partitioning processes; under common circumstances, their stable states are predicted by tendencies towards minimum free energies. But, living systems have *biochemical cycles,* where uphill steps are driven by a coupling to downhill steps. The persistence of the DEM for 3.5 billion years is because biology has become so capable of exploiting the food, energy, and matter out-of-equilibrium aspects of its environments. How might the DEM have arisen from prebiotic molecular processes? Below are some of the key questions.

Puzzles about the molecular origins of the DEM

- *What molecules were the first 'makers?'* Today's biological maker molecules are proteins and RNA, chain molecules that encode different functionalities as different monomer sequences. How did prebiotic polymers come to have sequence → function relationships? What simple molecular process started producing self-sustaining maker molecules?
- *How was molecule-making powered by external forces?* How did molecule–making come to outcompete molecule degradation and become so sustainable?
- *How did makers and catalysts become mobile, molecular-scale and editable?* An enormously transformative event in the prebiotic transition from chemistry to biology was the transition from catalysts that were immutable macroscale surfaces to microscale mobile editable proteins. Current thinking is that prebiotic reactions were first catalysed by mineral or clay surfaces (Wächtershäuser, 1988), or interfaces (Holden *et al.,* 2022), or in hot volcanic vents (Martin *et al.,* 2008). Such catalysts are macroscopic, geographically immovable, and fixed in their single-reaction catalysis under fixed conditions. But cells need whole biochemical pathways, where multiple reactions are strung together to achieve complex chemistry. Each step has a tailored catalyst that provides precisely the right acceleration of precisely that reaction, is mobile and small enough to fit inside a cell, and functions in the same water solvent as all the other requisite catalysts at room temperature. As a metaphor, consider the importance in the Industrial Revolution of steam engines, which replaced immovable energy sources of rivers and waterfalls by power that was mobile and tailorable to circumstances. How did prebiotic catalysis 'learn' to become untethered from rigid macroscopics to become flexible mobile microscopic biopolymers?
- *What was 'fitness' before there were cells?* Biological organisms are self-serving. This is captured in the multi-faceted notion of *fitness.* In contrast, molecules are not self-serving. How would molecules start becoming selected for or against?
- *Needles in haystacks and blind watchmakers: overcoming the infinitesimal probabilities.* Life's origin is often considered impossibly improbable, like finding a needle in a haystack, or finding a watch made by a blind watchmaker (Dawkins *et al.,* 1996). But those arguments are based on models that assume many improbable steps happen independently. There is a problem with those models. Life's originating events were surely not independent: they were correlated. The key questions are: (1) What was the nature of those correlations? and (2) In what

physical process does each step build on the advantages of the preceding steps to give cumulative long-term sustainability?

The case for proteins and the folding process

On the one hand, our view is that even the earliest life requires multiple components – functional molecules like proteins, informational molecules like RNA/DNA, encapsulation like lipids, and on-board energy like the ATP; see Supplementary Material SI.3 and Carter and Kraut (1974) and Frenkel-Pinter *et al.* (2020). On the other hand, our goal here is more modest, namely just to explain the roots of evolution-like dynamics, how molecules became makers, and how maker molecules developed sequence-to-function relationships.

In principle, the first sequence-to-function maker molecules could have been either RNA or proteins. The pros and cons of the *RNA world hypothesis,* that RNA came first, are discussed elsewhere (Joyce *et al.,* 1994; Atkins *et al.,* 2011; Robertson and Joyce, 2012; Joyce and Szostak, 2018; Wills and Carter, 2018). Here, we postulate that proteins came first, both for reasons discussed in those references and because of the need to first establish some form of propagation dynamics. Here is a short summary. (1) Proteins are most of a cell's mass, so the differential growth rates of cell evolution are largely a matter of differential protein production. (2) Proteins are today's main maker molecules, catalysing the reactions of cell growth. (3) Proteins are unique in having sequence → structure → function relationships. Most other polymers, including most RNAs, do not. Proteins achieve their actions, functions and mechanisms by virtue of their *native molecular structures.* The folding code is primarily a hydrophobic (H) and polar (P) code, which other linear biomolecules do not have. Consequently, while some RNA molecules do fold uniquely and are catalysts, they are driven by different forces. Proteins are compact because they are dominated by hydrophobic tertiary interactions, whereas RNA molecules tend to be stringy because they are dominated by secondary-structure interactions of hydrogen bonding and base stacking. Moreover, because hydrogen bonds and base stacking are relatively sequence-independent, where chain slippage leads to many local minima in free energy, RNA folding landscapes are bumpier and less funnelled than protein landscapes (Chen and Dill, 2000). Thus, even RNAs that actually have folded structures tend to have multiple ones, and those structures are only weakly specified by RNA sequences. (4) Proteins' unique folded states make proteins good catalysts. Folded proteins are miniature solids. Being a solid is exactly what is needed to catalyse chemical reactions, because catalyst atoms need to hold their places long enough to assist the reaction. (5) A 20 amino acid alphabet spans a range of chemistries, so they catalyse a range of reactions. For these purposes, RNA molecules are not as good as proteins. Even where a given reaction can be catalysed by either proteins or RNAs, proteins are often better (Plaxco and Gross, 2021). (6) While some RNA molecules can self-copy, those molecules would need to have very low error rates in order to persist (Eigen, 1971; Jeancolas *et al.,* 2020). The first copying machines would have to have had near-perfect fidelities. However, exact copying would be too brittle, for the same reasons we explained above that winner-takes-all (WTA) competitions are: without a way of generating diversity, exact copying is too prone to extinction in the face of environmental changes. In our view, prebiotic forces did not aim at self-copying; they aimed instead towards becoming autocatalytic *sets,* not strict autocatalysts. Variance is crucial. Progeny must not be identical to moms. The origins
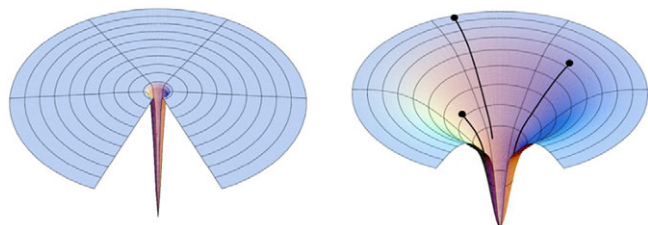
Fig. 2. Different landscapes of stochastic exploration: Golf courses *versus* Funnel Landscapes. Lateral directions are sampling degrees of freedom; the up-down direction is some measure of value (more value is downhill). (Left) *Blind Watchmaker, Needle-in-a-Haystack*: all states, except one, have no value. Success is nearly impossible. (Right) *Funnel Landscape:* From any starting point, there is often some direction that gives incremental advantage. And there are many routes for chaining together small advantages to more global advantage (black lines). Success is nearly inevitable. We believe evolution, once it gets going, is more funnel-like.

process must have some aspects of replication that are also to some degree unfaithful.

In terms of a dynamical process, protein folding has pertinent features. Protein folding entails a probabilistic needle-in-a-haystack search challenge through a disorder-to-order transformation. The folding search problem is now well understood in terms of *funnel-shaped energy landscapes* (Chan and Dill, 1991; Dill and Shortle, 1991; Wolynes *et al.,* 1995; Onuchic *et al.,* 1997; Wolynes, 1997; Dill *et al.,* 2007, 2008; Thirumalai *et al.,* 2010; Rollins and Dill, 2014; Nassar *et al.,* 2021). Fig. 2 compares a funnel landscape to a 'golf-course' landscape, which is premised on the assumption of uncorrelated independent events. 'Funnel' refers to the coarsest level of kinetic features, and not the potentially many finer-grained kinetic traps. Protein folding occurs so rapidly and towards such a unique ordered state because small random local steps combine together to lead effectively to the native state. In short, many proteins fold by rapidly finding needles in haystacks and creating complex watchmaker-like structures through small random correlated actions following combinatorially many microscopic routes via opportunistic chemical preferences. Protein folding gives both a metaphor for needle-in-a-haystack searching and a specific physical process, as described below, that could have become evolutionary dynamics.

Emergent autocatalysis from HP foldcats

Here is our hypothesis, first in overview, then in more detail. We postulate that prebiotic syntheses could produce short peptide

chains, some of which collapse into compact structures in water because of their hydrophobic content. A fraction of those collapsed chains will have exposed hydrophobic surfaces, active as a primitive catalytic site, slightly accelerating the binding and elongation of other peptides. Computer simulations show that this mechanism leads to autocatalytic sets. The premise that amino acids could be produced and could polymerise into short random peptide chains under plausible prebiotic conditions is well-established (Miller and Urey, 1959; Wächtershäuser, 1988; Botta and Bada, 2002; Johnson *et al.,* 2008; Lambert, 2008; Ikehara, 2014; Foden *et al.,* 2020; Frenkel-Pinter *et al.,* 2020; Muchowska *et al.,* 2020; Holden *et al.,* 2022; Krasnokutski *et al.,* 2022).

However, existing peptide synthesis experiments do not explain how chains could have become long enough to fold and function like proteins; how they could become catalysts and makers; how the process could become autocatalytic; how they could give non-random sequence → structure relationships; how catalysis became mobile; or what are the molecular origins of fitness. We address these below.

Fig. 3 illustrates the chain elongation challenge. Typical polymer syntheses give mostly only short chains that are not long enough to fold and function as today's proteins do. However, it has been found in computer modelling that some heteropolymers behave differently (Guseva *et al.,* 2017). Chains that have particular sequences of hydrophobic (H) and polar (P) types of monomers, called HP polymers, collapse in water into compact states due to the hydrophobic effect. Even some relatively short sequences can collapse. Here, we call those chains *foldamers*. Furthermore, a small fraction of foldamer sequences can act as primitive catalysts, described below.

HP chains can fold, catalyse and elongate

Fig. 4 shows that HP chain molecules have three general classes of behaviour in water, depending on their sequence of H and P monomers. (1) Some chains do not fold at all (think of the all-P sequence, for example). (2) Some HP sequences are foldamers, compact with hydrophobic cores. And (3) a fraction of HP foldamers happen to have surface patches that are concentrated in hydrophobic monomers; we call these surface regions 'landing pads', because these are regions that are sticky for other hydrophobic molecules floating in solution. Landing pads can be regions of catalysis. We call collapsed chains having landing pads *foldcats*, short for foldamer-catalysts.

These landing pads on foldcats could catalyse the covalent elongation of other 'client' HP sequences. The mechanism of this
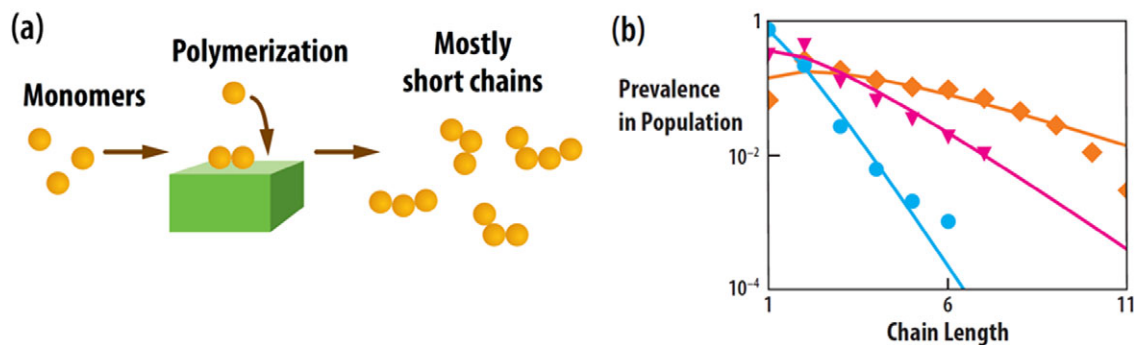

Fig. 3. Traditional polymerizations give mostly short chains, described by the Flory distribution. (*a*) A stationary catalyst polymerisation scheme. (*b*) Examples of the resulting Flory distribution, fit to experimental data: Orange (Kanavarioti *et al.,* 2001), pink (Ferris, 1999), and blue (Ding *et al.,* 1996). Populations rapidly diminish exponentially with chain length. This plot was reprinted with permission from Guseva *et al.* (2017).
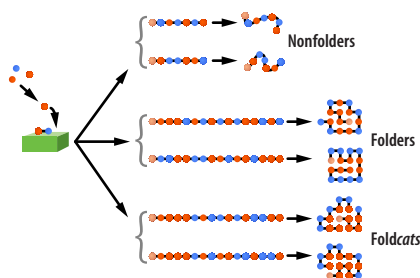
Fig. 4. Some HP chains can fold in water. HP chains are heteropolymers of hydrophobic (H, red) and polar (P, blue) monomers. Some sequences will not fold (nonfolders), while others will fold into compact states in water (*foldamers*), and a fraction of those sequences will have surface sites that can catalyse other reactions (*foldcats*).

catalysis process is shown in Fig. 5. Each foldcat sequence balls up, leaving a sticky spot (clustered H monomers) on its surface. A different peptide chain, call it a *client*, lands with its H monomers binding hydrophobically to the landing pad of the foldcat. A free H monomer from solution also lands on the landing pad. The spatial colocalization of the H monomer adjacent to the client chain can reduce the kinetic barrier to elongation of the client chain. The foldcat's job is to keep all the required pieces for elongation (the growing chain and a free monomer) in the same place. Peptide bond formation has a transition state barrier of 18 kcal mol^{-1} (Gindulyte *et al.,* 2006). Spatial localization of two reactants, often called *proximity effects* or *enhanced effective concentrations,* is known to accelerate covalent bond formation reactions by as much as 10^8 (Menger and Nome, 2019). For illustration, we have supposed only a binary code and only hydrophobicity-based landing pads. More realistically, a code will have more than two amino acid types, and more diverse interactions. The expectation that prebiotic peptides would have had both H and P amino acids is supported by the Miller–Urey experiment and recent variants of it (Miller and Urey, 1959; Botta and Bada, 2002; Johnson *et al.,* 2008). A proposed

minimal set would be GADV peptides (Ikehara, 2014), although there would be value in including cysteine (Foden *et al.,* 2020), and lysine or arginine for breadth of chemistry and control of aggregation.

This HP foldcat mechanism has recently been observed and explored in computer simulations (Guseva *et al.,* 2017). First, note that all these effects would likely have been almost negligibly small at first. Foldamers constitute only a fraction of all HP sequences; foldcats constitute an even smaller fraction; and colocalization-based rate enhancements are unlikely to be greater than a few kT in free energy (based on hydrophobicity estimates). But, it is not the *smallness* of populations or actions that matter. Rather, it is whether one step to the next entails some form of *systematic positive cooperativity*. What matters for origins (as well as for evolution in general) is whether some sub-population, even a very small one, is capable of some action – call it *emergent behaviour* – of positive feedback, so that it grows relative to other sub-populations, ultimately overcoming the relatively fixed forces of degradation. In general, a big challenge in origins-of-life research is that the initial seeding event is likely to be a very small signal in very large noise – precisely the sort of event for which devising a good experiment is difficult. Below, we describe how the HP foldcat mechanism predicts such emergent behaviours.

How the folding process leads to the evolution process

Here are the emergent behaviours of the HP folding and catalysis mechanism.

- *Emergence of makers, catalysts and molecular functionalities.* From the short random peptides that are plausibly synthesised prebiotically, the HP foldcat mechanism produces longer chains; see Fig. 6a. On average, longer HP chains are more stably folded and more protein-like (because they bury more hydrophobic surface). So, as long as amino acids are input, the HP foldamer mechanism pushes from peptides towards
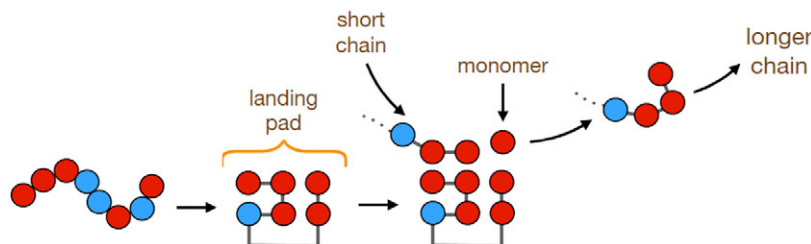

Fig. 5. HP foldcat chains can catalyse the elongation of other peptides. From the left: an HP sequence folds and exposes a hydrophobic surface ('landing pad'), a site on which a different chain can land and add a monomer to grow longer.
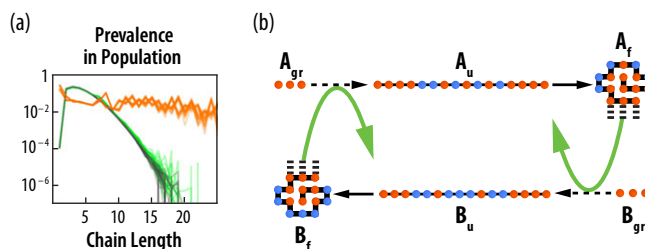

Fig. 6. The HP Foldcat mechanism: (*a*) grows longer chains and (*b*) populates an autocatalytic set. (*a*) Starting from random short HP molecules, chains elongate (orange) more than in the traditional Flory distribution (black) or the case of foldamers only with no catalysis (green). (*b*) Active, folded foldcat sequences (A_f and B_f) amplify the populations of other growing foldcats (A_{gr} and B_{gr}) while they are unfolded (A_u and B_u), leading to an *autocatalytic set*.

proteins, creating more catalytic power and functional diversity. These foldcats are makers that make makers. An alternative mechanism proposed for chain elongation is templated ligation, but it requires enzyme assistance (Tkachenko and Maslov, 2015; Kudella *et al.*, 2021).

- *Emergence of sequence-to-function relationships.* This mechanism amplifies the populations of foldamers and foldcats, simply because foldcats are a larger proportion of longer-chain sequences. Foldcats form an autocatalytic set; Fig. 6*b*. Such situations, where some sequences are populated selectively relative to other sequences based on their functionalities, are the basis for sequence-to-function relationships.

- *Emergence of programmable mobile molecular machines.* Presumably, the first prebiotic peptides were synthesised on macroscale catalysts, fixed in space and inflexible in their actions. But, the foldcat mechanism then produces its own catalysts, poor at first and better later. This untethers the peptide catalysis process from fixed spaces. Now, catalysts are at the microscale: they are mobile; and they are diverse and programmable by virtue of the sampling of sequence space. We regard this untethering, from macro to micro, from fixed to mobile, to have been a transformative step from prebiotic chemistry to biology.

- *Emergence of adaptation.* Arguably, evolution's central principle is that organisms adapt to environments. Evolution's great power of innovation and resourcefulness comes from its mutational search, competition, and fitness-based selection. The HP foldamer perspective posits that such adaptivity could have originated from a disorder-to-order process, in which chain molecules sample different sequences; molecules compete for limited resources; and winners are those that are more stable and get more resources.

- *What is 'fitness' among molecules? First, just persistence.* Darwinian evolution chooses winners and losers based on fitness ratcheting. What are winners and losers in a prebiotic world of molecules? HP chains persist in stably folded states for longer or shorter times, based on their sequences. Longer chains are more stable because they bury more hydrophobic residues upon folding, and because compactness limits access to chemical agents that hydrolyze proteins. In unruly environments of booms and busts, molecules that are more stable persist by scavenging the recycled monomers and peptides from molecules that are less stable.

- *Emergence of a tipping point from error catastrophes to success catastrophe.* Prebiotic molecules are subject to degradation. Error catastrophes are unavoidable in direct-replicator mechanisms (Eigen, 1971; Jeancolas *et al.*, 2020). Short peptides will hydrolyze to monomers. The origin of life was a tipping point from error catastrophes (where degradation dominates), to a 'success catastrophe', where maker molecules establish persistent populations. Beyond this point, evolution and growth then prevail over degradation. Three factors explain this tipping point in the HP foldamer model: *(1) Autocatalysis.* As noted above, peptides grow longer, more stably folded, and form an autocatalytic set. This contributes positive cooperativity towards self-sustainability. *(2) A driven machine.* Like a TV set that is 'plugged in', the HP foldamer mechanism is driven by a persistent input. The input is amino acids (and at early stages, also a catalyst of peptide synthesis). It does not matter that most product peptides fail and degrade; being 'plugged in' means that the system keeps pumping to push the chain lengths higher.
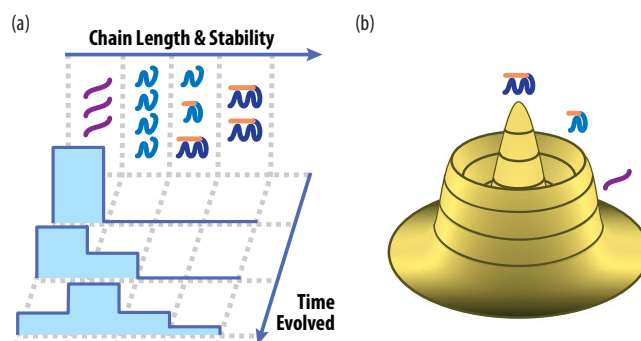

Fig. 7. The HP foldcat mechanism spontaneously grows populations of longer chains. (*a*) Longer proteins fold more stably and are more persistent to fluctuating environments. (*b*) A conceptual fitness landscape based on time of persistence of folding stability.

(3) Adaptivity. Autocatalysis and input power alone are not sufficient. Environments are unruly. Biology would not have survived without adaptability to changing environments. The combination of these factors contribute to a drive to ratchet up persistence over time; see Fig. 7.

To summarise, the HP foldamer mechanism explains how peptide synthesis and folding could result in the emergence of evolution-like propagation; see Table 1. But to be clear, we regard this not as the origin of life itself, but rather only as a precursor to it. Origins surely required much more than this: cell-like encapsulation, information and heritability, and more (some further discussion is given in the Supplementary Material).

Evidence supporting the HP foldcat mechanism

Although there is no direct experiment testing this foldcat mechanism, several of its components are supported by experiments. *HP chains can fold and function.* The binary HP code dominates protein folding (Lim and Sauer, 1989; Bowie *et al.*, 1990; Kamtekar *et al.*, 1993; Dill *et al.*, 1995; Dill and MacCallum, 2012; Koga *et al.*, 2020). But also, a biomolecule backbone is not required; HP peptoids can fold and function too (Lee *et al.*, 2005; Yoo and Kirshenbaum, 2008). *Some random peptide sequences can fold.* It is not an infinitely dilute space of sequences that can fold. As discussed in Guseva *et al.* (2017), for HP chains up to length 25, 2.3% fold to unique structures and 12.7% of those foldamers, or 0.3% of all sequences, have the foldcat catalytic surface. *Peptide syntheses occur naturally.* Even in interstellar space, 6–8-mer peptides have been found (Kebukawa *et al.*, 2022; Krasnokutski *et al.*, 2022). Sea spray

Table 1. Correspondence between properties of evolution and properties of origins, in the HP foldamer model described in the text

DEM	Origins
Moms	Protein molecules
… make moms	… make foldcats
Mutational search	Random sequences
Fitness	Persistence
Fitness gain	Folding stability
Degrees of freedom	Chain length

or air-water surfaces could catalyse small peptide formation (Griffith and Vaida, 2012; Deal *et al.,* 2021; Holden *et al.,* 2022). *Some short peptides can catalyse reactions* (Adamala and Szostak, 2013; Rufo *et al.,* 2014). *Hydrophobic patches are common on proteins* (Lijnzaad *et al.,* 1996; Tonddast-Navaei and Skolnick, 2015), which we call landing pads in the foldamer mechanism. *Some proteins are synthesised without ribosomes* (Finking and Marahiel, 2004; Miller and Gulick, 2016).

Outlook: from protein folding to evolution

We have posited that the origins of life could not have arisen without first a Darwin-like propagation mechanism. We believe function came before information, because we know of no driving force for the reverse. Rather than *genes using proteins to make new genes* (as in the Selfish Gene hypothesis Dawkins, 1976), our view is that *proteins use genes to make new proteins*. And, the foldcat mechanism indicates a way that the middleman – the gene – was simply not needed at first. This mechanism is based on solution physics – the oil–water and hydrogen-bonding forces of protein folding, the ability of miniature solids having different chemical moieties to catalyse reactions, and the ability of random syntheses to find and retain useful sequences based on their persistences. The foldcat mechanism addresses an important problem of origins research: it does not require a guiding hand of a researcher who chooses molecules, systems or processes. Instead, the foldcat mechanism is a disorder-to-order transition that bootstraps functional advantages that it finds from random search.

References

Adamala K and Szostak JW (2013) Competition between model protocells driven by an encapsulated catalyst. *Nature Chemistry* **5**(6), 495–501.

Armstrong RA and McGehee R (1980) Competitive exclusion. *The American Naturalist* **115**(2), 151–170.

Atkins JF, Gesteland RF and Cech T (eds) (2011) *RNA Worlds: From Life's Origins to Diversity in Gene Regulation.* New York: Cold Spring Harbor Laboratory Press.

Barabás G, D'Andrea R and Stump SM (2018) Chesson's coexistence theory. *Ecological Monographs* **88**(3), 277–303.

Beddington JR (1975) Mutual interference between parasites or predators and its effect on searching efficiency. *The Journal of Animal Ecology* **44**, 331–340.

Benner SA (2010) Defining life. *Astrobiology* **10**(10), 1021–1030.

Botta O and Bada JL (2002) Extraterrestrial organic compounds in meteorites. *Surveys in Geophysics* **23**, 411–467.

Bowie JU, Reidhaar-Olson JF, Lim WA and Sauer RT (1990) Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science* **247**(4948), 1306–1310.

Carter Jr CW and Kraut J (1974) A proposed model for interaction of polypeptides with RNA. *Proceedings of the National Academy of Sciences* **71**(2), 283–287.

Chan HS and Dill KA (1991) Polymer principles in protein structure and stability. *Annual Review of Biophysics and Biophysical Chemistry* **20**(1), 447–490.

Charlebois DA and Balázsi G (2016) Frequency-dependent selection: A diversifying force in microbial populations. *Molecular Systems Biology* **12**(8), 880.

Chen S-J and Dill KA (2000) RNA folding energy landscapes. *Proceedings of the National Academy of Sciences* **97**(2), 646–651.

Chesson P (1990) Macarthur's consumer-resource model. *Theoretical Population Biology* **37**(1), 26–38.

Chesson P (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics* **31**, 343–366.

Chesson P and Kuang JJ (2008) The interaction between predation and competition. *Nature* **456**(7219), 235–238.

Cleland CE and Chyba CF (2002) Defining 'life'. *Origins of Life and Evolution of the Biosphere* **32**(4), 387–393.

Cody GD (2004) Transition metal sulfides and the origins of metabolism. *Annual Review of Earth and Planetary Sciences* **32**, 569.

Dawkins R (1976). *The Selfish Gene.* New York: Oxford University Press.

Dawkins R. (1996) *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design.* New York: WW Norton & Company.

De Duve C (1995) The beginnings of life on earth. *American Scientist* **83**(5), 428–437.

De Duve C and De Neufville R (1991) *Blueprint for a Cell: The Nature and Origin of Life.* Burlington, NC: Carolina Biological Supply Company.

Deal AM, Rapf RJ and Vaida V (2021) Water–air interfaces as environments to address the water paradox in prebiotic chemistry: A physical chemistry perspective. *The Journal of Physical Chemistry A* **125**(23), 4929–4942.

Deamer D (2011) *First Life: Discovering the Connections between Stars, Cells, and How Life Began.* Berkley, CA: University of California Press.

DeAngelis DL, Goldstein R and O'Neill RV (1975) A model for tropic interaction. *Ecology* **56**(4), 881–892.

Dill KA, Bromberg S, Yue K, Chan HS, Ftebig KM, Yee DP and Thomas PD (1995) Principles of protein folding—A perspective from simple exact models. *Protein Science* **4**(4), 561–602.

Dill KA and MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* **338**(6110), 1042–1046.

Dill KA, Ozkan SB, Shell MS and Weikl TR (2008) The protein folding problem. *Annual Review of Biophysics* **37**, 289.

Dill KA, Ozkan SB, Weikl TR, Chodera JD and Voelz VA (2007) The protein folding problem: When will it be solved? *Current Opinion in Structural Biology* **17**(3), 342–346.

Dill KA and Shortle D (1991) Denatured states of proteins. *Annual Review of Biochemistry* **60**(1), 795–825.

Ding PZ, Kawamura K and Ferris JP (1996) Oligomerization of uridine phosphorimidazolides on montmorillonite: A model for the prebiotic synthesis of RNA on minerals. *Origins of Life and Evolution of the Biosphere* **26**, 151–171.

Doebeli M, Jaque EC and Ispolatov Y (2021) Boom-bust population dynamics increase diversity in evolving competitive communities. *Communications Biology* **4**(1), 1–8.

Dyson FJ (1982) A model for the origin of life. *Journal of Molecular Evolution* **18**(5), 344–350.

Dyson F (1999) *Origins of Life.* Cambridge: Cambridge University Press.

Eigen M (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**(10), 465–523.

Ferris JP (1999) Prebiotic synthesis on minerals: Bridging the prebiotic and RNA worlds. *The Biological Bulletin* **196**(3), 311–314.

Finking R and Marahiel MA (2004) Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology* **58**, 453.

Fisher R (1930) *The Genetical Theory of Natural Selection.* Oxford: Clarendon.

Foden CS, Islam S, Fernández-Garca C, Maugeri L, Sheppard TD and Powner MW (2020) Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**(6518), 865–869.

Frenkel-Pinter M, Haynes JW, Mohyeldin AM, Sargon AB, Petrov AS, Krishnamurthy R, Hud NV, Williams LD and Leman LJ (2020) Mutually

stabilizing interactions between proto-peptides and RNA. *Nature Communications* **11**(1), 3137.

Gause G (1934) *The Struggle for Existence*. Baltimore, MA: Williams and Wilkins.

Gilbert W (1986) Origin of life: The RNA world. *Nature* **319**(6055), 618–618.

Gindulyte A, Bashan A, Agmon I, Massa L, Yonath A and Karle J (2006) The transition state for formation of the peptide bond in the ribosome. *Proceedings of the National Academy of Sciences* **103**(36), 13327–13332.

Goyal A, Dubinkina V and Maslov S (2018) Multiple stable states in microbial communities explained by the stable marriage problem. *The ISME Journal* **12** (12), 2823–2834.

Griffith EC and Vaida V (2012) In situ observation of peptide bond formation at the water–air interface. *Proceedings of the National Academy of Sciences* **109**(39), 15697–15701.

Guseva E, Zuckermann RN and Dill KA (2017) Foldamer hypothesis for the growth and sequence differentiation of prebiotic polymers. *Proceedings of the National Academy of Sciences* **114**(36), E7460–E7468.

Holden DT, Morato NM and Cooks RG (2022) Aqueous microdroplets enable abiotic synthesis and chain extension of unique peptide isomers from free amino acids. *Proceedings of the National Academy of Sciences* **119**(42), e2212642119.

Hordijk W (2019) A history of autocatalytic sets. *Biological Theory* **14**(4), 224–246.

Hordijk W and Steel M (2014) Conditions for evolvability of autocatalytic sets: A formal example and analysis. *Origins of Life and Evolution of Biospheres* **44** (2), 111–124.

Hordijk W, Steel M and Kauffman S (2022) Autocatalytic sets arising in a combinatorial model of chemical evolution. *Life* **12**(11), 1703.

Hsu S-B, Hubbell S and Waltman P (1977) A mathematical theory for single-nutrient competition in continuous cultures of micro-organisms. *SIAM Journal on Applied Mathematics* **32**(2), 366–383.

Hutchinson GE (1961) The paradox of the plankton. *The American Naturalist* **95**(882), 137–145.

Ikehara K (2014) [GADV]-protein world hypothesis on the origin of life. *Origins of Life and Evolution of Biospheres* **44**, 299–302.

Jain S and Krishna S (2002) Large extinctions in an evolutionary model: The role of innovation and keystone species. *Proceedings of the National Academy of Sciences* **99**(4), 2055–2060.

Jeancolas C, Malaterre C and Nghe P (2020) Thresholds in origin of life scenarios. *Iscience* **23**(11), 101756.

Johnson AP, Cleaves HJ, Dworkin JP, Glavin DP, Lazcano A and Bada JL (2008) The miller volcanic spark discharge experiment. *Science* **322**(5900), 404–404.

Jordan SF, Ioannou I, Rammu H, Halpern A, Bogart LK, Ahn M, Vasiliadou R, Christodoulou J, Maréchal A and Lane N (2021) Spontaneous assembly of redox-active iron-sulfur clusters at low concentrations of cysteine. *Nature Communications* **12**(1), 1–14.

Joyce G, Deamer DW and Fleischaker G (1994) *Foreward to Origins of Life: The Central Concepts*. Boston, MA: Jones and Bartlett Publishers.

Joyce GF and Szostak JW (2018) Protocells and RNA self-replication. *Cold Spring Harbor Perspectives in Biology* **10**(9), a034801.

Kamtekar S, Schiffer JM, Xiong H, Babik JM and Hecht MH (1993) Protein design by binary patterning of polar and nonpolar amino acids. *Science* **262** (5140), 1680–1685.

Kanavarioti A, Monnard P-A and Deamer DW (2001) Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology* **1**(3), 271–281.

Kauffman SA (1971) Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *Journal of Cybernetics* **1**, 71–96.

Kauffman SA (1986) Autocatalytic sets of proteins. *Journal of Theoretical Biology* **119**(1), 1–24.

Kauffman SA, et al. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford: Oxford University Press.

Kebukawa Y, Asano S, Tani A, Yoda I and Kobayashi K (2022) Gamma-ray-induced amino acid formation in aqueous small bodies in the early solar system. *ACS Central Science* **8**, 1664–1671.

Kocher, Charles D., and Ken A. Dill (2023) Darwinian evolution as a dynamical principle. *Proceedings of the National Academy of Sciences* **120**.11: e2218390120.

Koga R, Yamamoto M, Kosugi T, Kobayashi N, Sugiki T, Fujiwara T and Koga N (2020) Robust folding of a de novo designed ideal protein even with most of the core mutated to valine. *Proceedings of the National Academy of Sciences* **117**(49), 31149–31156.

Krasnokutski S, Chuang K-J, Jäger C, Ueberschaar N and Henning T (2022) A pathway to peptides in space through the condensation of atomic carbon. *Nature Astronomy* **6**(3), 381–386.

Kudella PW, Tkachenko AV, Salditt A, Maslov S and Braun D (2021) Structured sequences emerge from random pool when replicated by templated ligation. *Proceedings of the National Academy of Sciences* **118**(8), e2018830118.

Lambert J-F (2008) Adsorption and polymerization of amino acids on mineral surfaces: A review. *Origins of Life and Evolution of Biospheres* **38**(3), 211–242.

Lee B-C, Zuckermann RN and Dill KA (2005) Folding a nonbiological polymer into a compact multihelical structure. *Journal of the American Chemical Society* **127**(31), 10999–11009.

Lifson S (1997) On the crucial stages in the origin of animate matter. *Journal of Molecular Evolution* **44**(1), 1–8.

Lijnzaad P, Berendsen HJ and Argos P (1996) Hydrophobic patches on the surfaces of protein structures. *Proteins: Structure, Function, and Bioinformatics* **25**(3), 389–397.

Lim WA and Sauer RT (1989) Alternative packing arrangements in the hydrophobic core of λ repressor. *Nature* **339**(6219), 31–36.

MacArthur R (1970) Species packing and competitive equilibrium for many species. *Theoretical Population Biology* **1**(1), 1–11.

Machery E (2012) Why I stopped worrying about the definition of life… And why you should as well. *Synthese* **185**(1), 145–164.

Martin W, Baross J, Kelley D and Russell MJ (2008) Hydrothermal vents and the origin of life. *Nature Reviews Microbiology* **6**(11), 805–814.

Matsuo M and Kurihara K (2021) Proliferating coacervate droplets as the missing link between chemistry and biology in the origins of life. *Nature Communications* **12**(1), 1–13.

Maury CPJ (2009) Self-propagating β-sheet polypeptide structures as prebiotic informational molecular entities: The amyloid world. *Origins of Life and Evolution of Biospheres* **39**(2):141–150.

Maury CPJ (2015) Origin of life. Primordial genetics: Information transfer in a pre-RNA world based on self-replicating beta-sheet amyloid conformers. *Journal of Theoretical Biology* **382**, 292–297.

Maury CPJ (2018) Amyloid and the origin of life: Self-replicating catalytic amyloids as prebiotic informational and protometabolic entities. *Cellular and Molecular Life Sciences* **75**(9), 1499–1507.

Menger FM and Nome F (2019) Interaction vs. preorganization in enzyme catalysis. A dispute that calls for resolution. *ACS Chemical Biology* **14**(7), 1386–1392.

Miller BR and Gulick AM (2016) Structural biology of nonribosomal peptide synthetases. In *Nonribosomal Peptide and Polyketide Biosynthesis*. New York, NY: Springer, pp. 3–29.

Miller SL and Urey HC (1959) Organic compound synthesis on the primitive earth: Several questions about the origin of life have been answered, but much remains to be studied. *Science* **130**(3370), 245–251.

Muchowska KB, Varma SJ and Moran J (2020) Nonenzymatic metabolic reactions and life's origins. *Chemical Reviews* **120**(15), 7708–7744.

Nassar R, Dignon GL, Razban RM and Dill KA (2021) The protein folding problem: The role of theory. *Journal of Molecular Biology* **433**(20), 167126.

Novak M and Stouffer DB (2021) Systematic bias in studies of consumer functional responses. *Ecology Letters* **24**(3), 580–593.

Onuchic JN, Luthey-Schulten Z and Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry* **48**(1), 545–600.

Plaxco KW and Gross M (2021) *Astrobiology: An Introduction*. Baltimore, MD: Johns Hopkins University Press.

Popa R (2004) *Between Necessity and Probability: Searching for the Definition and Origin of Life*. Berlin: Springer Science & Business Media.

Pross A (2011) Toward a general theory of evolution: Extending Darwinian theory to inanimate matter. *Journal of Systems Chemistry* **2**(1), 1–14.

Pross A (2016) *What Is Life? How Chemistry Becomes Biology*. Oxford: Oxford University Press.

Robertson MP and Joyce GF (2012) The origins of the RNA world. *Cold Spring Harbor Perspectives in Biology* **4**(5), a003608.

Rollins GC and Dill KA (2014) General mechanism of two-state protein folding kinetics. *Journal of the American Chemical Society* **136**(32), 11420–11427.

Rufo CM, **Moroz YS**, **Moroz OV**, **Stöhr J**, **Smith TA**, **Hu X**, **DeGrado WF and Korendovych IV** (2014) Short peptides self-assemble to produce catalytic amyloids. *Nature Chemistry* **6**(4), 303–309.

Schrodinger E (1944) *What Is Life?* Cambridge: Cambridge University Press.

Segré D, **Ben-Eli D**, **Deamer DW and Lancet D** (2001) The lipid world. *Origins of Life and Evolution of the Biosphere* **31**(1), 119–145.

Shapiro R (2006) Small molecule interactions were central to the origin of life. *The Quarterly Review of Biology* **81**(2), 105–126.

Stouffer DB and Novak M (2021) Hidden layers of density dependence in consumer feeding rates. *Ecology Letters* **24**(3), 520–532.

Thirumalai D, **O'Brien EP**, **Morrison G and Hyeon C** (2010) Theoretical perspectives on protein folding. *Annual Review of Biophysics* **39**(1), 159–183.

Tilman D (1982) *Resource Competition and Community Structure. Monographs in Population Biology*. Princeton, NJ: Princeton University Press.

Tkachenko AV and Maslov S (2015) Spontaneous emergence of autocatalytic information-coding polymers. *The Journal of Chemical Physics* **143**(4), 07B612_1.

Tonddast-Navaei S and Skolnick J (2015) Are protein-protein interfaces special regions on a protein's surface? *The Journal of Chemical Physics* **143** (24), 12B631_1.

van Opheusden JH, **Hemerik L**, **van Opheusden M and van der Werf W** (2015) Competition for resources: Complicated dynamics in the simple Tilman model. *Springerplus* **4**(1), 1–31.

Volterra V (1928) Variations and fluctuations of the number of individuals in animal species living together. *ICES Journal of Marine Science* **3**(1), 3–51.

Wächtershäuser G (1988) Before enzymes and templates: Theory of surface metabolism. *Microbiological Reviews* **52**(4), 452–484.

Wang Z, **Fridman Y**, **Maslov S and Goyal A** (2022) Fine-scale diversity of microbial communities due to satellite niches in boom-and-bust environments. *PLOS Computational Biology*, **18**, e1010244

Wills PR and Carter CW (2018) Insuperable problems of the genetic code initially emerging in an RNA world. *Biosystems* **164**, 155–166.

Wolynes PG (1997) Folding funnels and energy landscapes of larger proteins within the capillarity approximation. *Proceedings of the National Academy of Sciences* **94**(12), 6170–6175.

Wolynes PG, **Onuchic JN and Thirumalai D** (1995) Navigating the folding routes. *Science* **267**(5204), 1619–1620.

Yoo B and Kirshenbaum K (2008) Peptoid architectures: Elaboration, actuation, and application. *Current Opinion in Chemical Biology* **12**(6), 714–721.